ISSN: 2348-2079

MAY-2015

International Journal of Intellectual Advancements and Research in Engineering Computations

# EFFICIENT AND PRIVACY-AWARE DATA AGGREGATION IN MOBILE SENSING

[1]R. Sangeetha, [2]K. Mubarak Ali.

## ABSTRACT

The proliferation and ever-increasing capabilities of mobile devices such as smart phones give rise to a variety of mobile sensing applications. This paper studies how a untrusted aggregator in mobile sensing can periodically obtain desired statistics over the data contributed by multiple mobile users, without compromising the privacy of each user. Although there are some existing works in this area, they either require bidirectional communications between the aggregator and mobile users in every aggregation period, or have high-computation overhead and cannot support large plaintext spaces. Also, they do not consider the Min aggregate, which is quite useful in mobile sensing. To address these problems, we propose an efficient protocol to obtain the Sum aggregate, which employs an additive homomorphic encryption and a novel key management technique to support large plaintext space. We also extend the sum aggregation protocol to obtain the Min aggregate of time-series data. To deal with dynamic joins and leaves of mobile users, we propose a scheme that utilizes the redundancy in security to reduce the communication cost for each join and leave. Evaluations show that our protocols are orders of magnitude faster than existing solutions and it has much lower communication overhead.

## INTRODUCTION

A sensor network consists of large number of sensors deployed in a region for the purpose of event monitoring or detection. The sensors are pre-programmed to listen for specific events. For example, a sensor network deployed in a high security region might be programmed to detect infrared heat signals to indicate an intruder. Figure 1 shows a typical sensor network deployment. Each node in a sensor network is responsible for observing and reporting various dynamic properties of their surroundings in a time critical manner. These mobile and miniaturized information devices are equipped with embedded processors, wireless communication circuitry, information storage capability, smart sensors and actuators. These sensor nodes networked in an ad hoc way, with little or no fixed network support, to provide the surveillance and targeting information for dynamic control. Sensor devices are mobile, subject to failure, deployed spontaneously and repositioned for more accurate surveillance. Despite these dynamic changes in configuration of the sensor network, critical real-time information must still be disseminated dynamically from mobile sensor data sources through the self-organizing network infrastructure to the components that control dynamic re-planning and re-optimization of the theatre of operation based on newly available information.

With large number of sensor devices being quickly and flexibly deployed in most impromptu networks, each sensor device must be autonomous and capable of organizing itself in the overall community of sensors to perform coordinated activities with global objectives. When spontaneously placed together in an environment, these sensor nodes should immediately know about the capabilities and functions of other sensor Nodes and work together as a community system to perform co-operative tasks and networking functionalities. Sensor networks need to be self-organizing since they

**Author for Correspondence:**

[1]Final year ME, Al-Ameen Engineering College, Erode, Tamilnadu, India.

[2]Assistant Professor/CSE, Al-Ameen Engineering College, Erode, Tamilnadu, India.

are often formed spontaneously from large number of mixed types of nodes and may undergo frequent configuration changes. Some sensor nodes may provide networking and system services and resources to other sensor nodes. Others may detect the presence of these nodes and request services from them.

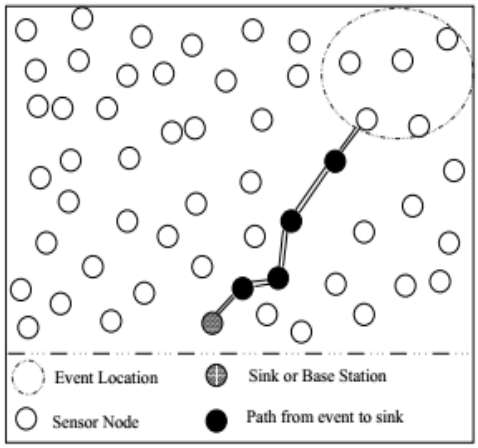


**Figure 1. A Typical Sensor Network Environment**

The characteristics of sensor nodes necessary for creating self-organizing sensor networks are agility, self awareness, self-configurability and autonomy. Sensor nodes with these features will have capabilities for self assembling impromptu networks that are incrementally extensible and dynamically adaptable to device failure and degradation, mobility of sensor nodes and changes in task and network requirements. Nodes are aware of their own capabilities and those of other nodes around them which may provide the networking and system services or resources that they need. Although nodes are autonomous, they may cooperate with one another to disseminate information or assist each other in adapting to changes in the network configuration. An impromptu community of these nodes may cooperate to provide continual coordinated services while some nodes may be newly deployed or removed from the spontaneous community.

Nodes will act in response to environmental events and relay collected and possibly aggregated information through the multi-hop wireless network in accordance with desired system functionality. The inherently dynamic and distributed behavior of these networks, coupled with inherent physical limitations such as small instruction and data memory, constrained energy resources, short communication radii and a low bandwidth medium in which to communicate, make developing communication protocols difficult. Using these sensors as a basis for development, the software architecture and communication stack residing on these devices are built taking into consideration the prolific research in the areas of ad-hoc networking, data aggregation, cluster formation, distributed services, group formation, channel contention and power conservation. An event is an abstraction, identifying anything from a set of sensor readings, to the nodes processing capabilities. For the purpose of the simulation studies in this project, events are assumed to be localized phenomenon, occurring in a fixed region of space. This assumption will hold for a wide variety of sensor-net applications, since many external events are localized themselves.

## RELATED WORK

Many works have addressed various security and privacy issues in mobile sensing networks and systems (e.g., [10], [2]), but they do not consider data

aggregation. There are a lot of existing works on security and privacy-preserving data aggregation, but most of them assume a trusted aggregator and cannot protect user privacy against untrusted aggregators. Yang et al. proposed an encryption scheme that allows an untrusted aggregator to obtain the sum of multiple users's data without knowing any specific user's data. Their scheme requires expensive rekeying operations to support multiple time steps and thus may not work for time-series data. Shi et al. proposed a privacy-preserving data aggregation scheme based on data slicing and mixing techniques. Their scheme is not designed for time-series data. It may not work well for time-series data, since each user may need to select a new set of peers in each aggregation interval due to mobility. Besides, their scheme for nonadditive aggregates requires multiple rounds of bidirectional communications between the aggregator and mobile users which means long delays. In contrast, our scheme obtains those aggregates with just one round of unidirectional communication from users to the aggregator.

To achieve privacy-preserving sum aggregation of time series data, Rastogi and Nath [6] designed an encryption scheme based on threshold Paillier cryptosystem, where the decryption key is divided into portions and distributed to the users. The aggregator collects the ciphertexts of users, multiplies them together and sends the aggregate ciphertext to all users. Each user decrypts a share of the sum aggregate. The aggregator collects all the shares and gets the final sum. Their scheme requires an extra round of interaction between the aggregator and users in every aggregation period. Erkin and Tsudik [12] also proposed an aggregation scheme based on Paillier cryptosystem, but it requires communications between every pair of users in every aggregation period.

Based on an efficient additive homomorphic encryption scheme, Rieffel et al. [9] proposed a construction that does not require an extra round of interaction between the aggregator and the users. In their scheme, the computation and storage cost is roughly equal to the number of colluding users that the system can tolerate. Thus, their scheme has high overhead to achieve good resistance to collusion, especially when the system is large and a large number of users collude. In contrast, our scheme tolerates a high fraction of colluding users with very small cost even when the system is large. Acs and Castelluccia [13] also proposed a scheme based on additive homomorphic encryption, but in their scheme each node shares a pairwise key with any other node.

Shi et al. [7] proposed a construction for sum aggregation based on the assumption that the Decisional Diffie-Hellman problem is hard over finite cyclic groups. In their construction, each user sends her ciphertext to the aggregator and no communication is needed from the aggregator to the users. To decrypt the sum, their construction needs to traverse the possible plaintext space of sum and thus, it is not efficient for a large system with large plaintext spaces. Chan et al. [8] extended the construction with a binary interval tree technique, but their scheme still has the limitation in plaintext spaces. Jawurek and Kerschbaum [14] proposed a scheme that provides differential privacy for sum. Our aggregation protocol for sum can be used as a building block of their scheme to improve the computational efficiency. Also, existing works [3] do not consider the Min of time series data.

## DATA AGGREGATION IN SENSOR NETWORKS

The source information for data aggregators may originate from public records and criminal databases; the information is packaged into aggregate reports and then sold to businesses, as well as to local, state and federal government agencies. This information can also be useful for marketing purposes. Many data brokers' activities fall under the Fair Credit Reporting Act (FCRA) which regulates consumer reporting agencies. The agencies then gather and package personal information into consumer reports that are sold to creditors, employers, insurers and other businesses.

Various reports of information are provided by database aggregators. Individuals may request their own consumer reports which contain basic biographical information such as name, date of birth, current address and phone number. Employee background check reports, which contain highly detailed information such as past addresses and length of residence, professional licenses and

criminal history, may be requested by eligible and qualified third parties. Not only can this data be used in employee background checks, but it may also be used to make decisions about insurance coverage, pricing and law enforcement. Privacy activists argue that database aggregators can provide erroneous information.

**NO DATA AGGREGATION**

In No Data Aggregation scheme, sensor devices are unaware of other neighboring nodes. Each sensor upon detecting an event attempts to send the amount of information collected, however small it may be, to the end nodes (sink). Sensor devices do not apply any data aggregation technique and simply forward the data packets toward the sink node. As we can clearly see, such a scheme suffers from high packet dropping rate and low bandwidth due to congestion in the network. Additionally, it also suffers from energy limitations as each device attempts to send packets received from multiple destinations irrespective of the importance of the data being transmitted. Furthermore, the total amount of information received at the sink nodes would be less due to several packets getting dropped. However such schemes may become useful under scenarios like battlefield or military surveillance where events may move at a very fast rate.

**PERFECT DATA AGGREGATION**

Hypothetical scenario, each sensor device is assumed to know the best data aggregator. In other words, the sensor device which would send the most critical information about a particular event is predefined. Such an environment is highly desirable since the sink nodes get the most critical, complete information about the events and such a scheme results in high bandwidth, improved response time and adheres to the energy constraints. However, in an environment which is highly dynamic in nature and with unpredictable traffic patterns achieving such an environment is almost impossible.

**IN-NETWORK DATA AGGREGATION**

This scheme is highly suitable for environments where events have localized phenomenon, occurring in a fixed region of space. Such environments will hold for a wide variety of sensor network applications, since many external events are localized themselves. In this scheme, the sensor network environment is divided into pre-defined set of grids or regions. Each region or grid is responsible for observing and reporting events that occur inside the region to the sink nodes. Also each sensor device inside the region sends data to other sensor devices only inside the region. Only one sensor, the data aggregator, sends the critical information received either from other sensor devices or by itself to the sink nodes.

A typical in-network data aggregation scheme. As we see in the figure, all sensor devices inside the region detect the event. The corresponding signal strengths detected by each sensor are shown in the figure. Now each sensor transmits its signal strength only to its neighbors'. If the neighbor has strength more than the sender, the sender decides to remain silent and stops transmitting packets. Otherwise, it waits for packets from other sensors and after receiving packets from all its neighbors', if the sender has the highest signal strength, it will then become the data aggregator and all other sensor devices stop detecting the event and helps only in routing the packet to the sink nodes.

**GRID-BASED DATA AGGREGATION**

Grid-based Data Aggregation is highly suitable for mobile environments where the time duration of an event at a particular place is very small. Such scenarios will hold for a variety of sensor network applications like military surveillance, weather forecasting, etc. As seen in the previous scheme, the sensor network environment is divided into pre-defined set of grids or regions. Each region or grid is responsible for observing and reporting events that occur inside the region to the sink nodes. In addition, one sensor device based on geographical position with respect to either the sink or the center of the grid is chosen as data aggregator. All other sensors inside the gird are aware of this information. During event detection, all other sensors are supposed to send the event information to this data aggregator. The data aggregator after collecting data from other sensors sends only the critical information to the sink node.

A typical Grid-based data aggregation scheme. As seen in the figure, during event detection, all sensors send data to the aggregator. After

collecting all data from other sensors, the aggregator sends only the critical information to the sink nodes. Grid-based data aggregation adapts well to dynamic changes in the network topology and event mobility. If the event is highly mobile in nature, we see that many packets are exchanged between the sensors inside the grid. But, once the packets reach the aggregator, we see that only the most important information is sent to the sink nodes. Thus, Grid-based scheme reduces the traffic in such environments and makes sure the critical information is transmitted to the end nodes interested in the data. It also increases the throughput in such environments. However Grid-based scheme performs worse in environments where events are highly localized and mostly immobile in nature. The data packets exchanged between the aggregator and other sensors inside the grid falls in the critical path. This increases the end-to-end response time.

## SECURE DATA AGGREGATION IN WSN

A Wireless Sensor Network (WSN) typically consists of a sink node sometimes referred to as a Base Station and a number of small wireless sensor nodes. The base station is assumed to be secure with unlimited available energy while the sensor nodes are assumed to be unsecured with limited available energy. The sensor nodes monitor a geographical area and collect sensory information. Sensory information is communicated to the Base Station through Wireless hop by hop transmissions. To conserve energy this information is aggregated at intermediate sensor nodes by applying a suitable aggregation function on the received data. Aggregation reduces the amount of network traffic which helps to reduce energy consumption on sensor nodes. It however complicates the already existing security challenges for wireless sensor networks and requires new security techniques tailored specifically for this scenario. Providing security to aggregate data in Wireless Sensor Networks is known as Secure Data Aggregation in WSN were the first few works discussing techniques for secure data aggregation in Wireless Sensor Networks. Two main security challenges in secure data aggregation are confidentiality and integrity of data. While traditionally encryption is used to provide end to end confidentiality in Wireless Sensor Network (WSN), the aggregators in a secure data aggregation scenario need to decrypt the encrypted data to perform aggregation. This exposes the plaintext at the aggregators, making.

## PRIVACY-AWARE DATA AGGREGATION IN MOBILE SENSING

Mobile devices such as smart phones are gaining an ever-increasing popularity. Most smart phones are equipped with a rich set of embedded sensors such as camera, microphone, GPS, accelerometer, ambient light sensor, gyroscope and so on. The data generated by these sensors provide opportunities to make sophisticated inferences about not only people but also their surrounding and thus can help improve people's health as well as life. This enables various mobile sensing applications such as environmental monitoring [1], traffic monitoring [2], healthcare.

In many scenarios, aggregation statistics need to be periodically computed from a stream of data contributed by mobile users [4], to identify some phenomena or track some important patterns. For example, the average amounts of daily exercise (which can be measured by motion sensors [5]) that people do can be used to infer public health conditions. The average or maximum level of air pollution and pollen concentration in an area may be useful for people to plan their outdoor activities. Other statistics of interests include the lowest gasoline price in a city, the highest moving speed of road traffic during rush hour and so on.

Although aggregation statistics computed from time series data are very useful, in many scenarios, the data from users are privacy-sensitive and users do not trust any single third-party aggregator to see their data values. For instance, to monitor the propagation of a new flu, the aggregator will count the number of users infected by this flu. A user may not want to directly provide her true status if she is not sure whether the information will be abused by the aggregator. Accordingly, systems that collect users' true data values and compute aggregate statistics over them may not meet users' privacy

requirement [4]. Thus, an important challenge is how to protect the users' privacy in mobile sensing, especially when the aggregator is untrusted.

Most previous works on sensor data aggregation assume a trusted aggregator and hence cannot protect user privacy against an untrusted aggregator in mobile sensing applications. Several recent works [7] consider the aggregation of time-series data in the presence of an untrusted aggregator. To protect user privacy, they design encryption schemes in which the aggregator can only decrypt the sum of all users' data but nothing else. Rastogi and Nath [6] use threshold Paillier cryptosystem build such an encryption scheme. To decrypt the sum, their scheme needs an extra round of interaction between the aggregator and all users in every aggregation period, which means high communication cost and long delay. Moreover, it requires all users to be online until decryption is completed, which may not be practical in many mobile sensing scenarios due to user mobility and the heterogeneity of user connectivity. Rieffel et al. [9] propose a construction that does not require bidirectional communications between the aggregator and the users, but it has high computation and storage cost to deal with collusions in a large system.

Shi et al. [7], [8] also propose a construction for sum aggregation, which does not need the extra round of interaction. The decryption in their construction needs to traverse the possible plaintext space of the aggregated value, which is very expensive for a large system with large plaintext space. In mobile sensing, the plaintext space of some application can be large. For example, carbon dioxide levels can range from 350 ppm outdoors to over 10,000 ppm in industrial workplaces [11]. Hence, in applications that continuously monitor the carbon dioxide levels that people are exposed to in their daily life, the plaintext space can reach 104. Under this plaintext space, for a large system with one million users, the construction in [7] requires 30 seconds to decrypt the sum on a modern 64-bit desktop PC. Its computation overhead is too high for an aggregator to run real-time monitoring applications with short aggregation intervals and to collect multiple aggregate statistics simultaneously. Moreover, none of these existing schemes considers the Min aggregate of time series data, which is also important in many mobile sensing applications.

In this paper, we propose a new protocol for mobile sensing to obtain the sum aggregate of time-series data in the presence of an untrusted aggregator. Our protocol employs an additive homomorphic encryption and a novel key management scheme based on efficient HMAC to ensure that the aggregator can only obtain the sum of all users' data, without knowing individual user's data or intermediate result. In our protocol, each user only needs to compute a very small number of HMACs to encrypt her data. Hence, the computation cost is very low and the protocol can scale to large systems with large plaintext spaces, resource constrained devices and high aggregation loads. Another nice property of our protocol is that it only requires a single round of user-to-aggregator communication.

Based on the sum aggregation protocol, we propose a protocol to obtain the Min aggregate. To our best knowledge, this is the first privacy-preserving solution to obtain the Min of time-series data in mobile sensing with just one round of user to-aggregator communication. Our protocols for Sum and Min can be easily adapted to derive many other aggregate statistics such as Count, Average and Max. Since users may frequently join and leave in mobile sensing, we also propose a scheme that employs the redundancy in security to reduce the communication cost of dealing with dynamic joins and leaves.

## PROBLEM STATEMENT

Sensor data aggregation assumes a trusted aggregator and hence cannot protect user privacy against an untrusted aggregator in mobile sensing applications. Several recent works consider the aggregation of time-series data in the presence of an untrusted aggregator. To protect user privacy, they design encryption schemes in which the aggregator can only decrypt the sum of all users' data but nothing else. Use threshold Paillier cryptosystem to build such an encryption scheme. To decrypt the sum, their scheme needs an extra round of interaction between the aggregator and all users in every aggregation period, which means high communication cost and long delay. Moreover, it requires all users to be online until decryption is

completed, which may not be practical in many mobile sensing scenarios due to user mobility and the heterogeneity of user connectivity. The following problems are identified by the existing system,

- Cannot protect user privacy against untrusted aggregators.
- Existing works do not consider the Min of time-series data.

## PRIVACY PRESERVED DATA AGGREGATION FOR MOBILE SENSING

A new protocol for mobile sensing to obtain the sum aggregate of time-series data in the presence of an untrusted aggregator. Our protocol employs an additive homomorphic encryption and a novel key management scheme based on efficient HMAC to ensure that the aggregator can only obtain the sum of all users' data, without knowing individual user's data or intermediate result. In our protocol, each user only needs to compute a very small number of HMACs to encrypt her data. Hence, the computation cost is very low and the protocol can scale to large systems with large plaintext spaces, resource constrained devices and high aggregation loads. Another nice property of our protocol is that it only requires a single round of user-to-aggregator communication.

Based on the sum aggregation protocol, we propose a protocol to obtain the Min aggregate. To our best knowledge, this is the first privacy-preserving solution to obtain the Min of time-series data in mobile sensing with just one round of user-to-aggregator communication. Our protocols for Sum and Min can be easily adapted to derive many other aggregate statistics such as Count, Average and Max. Since users may frequently join and leave in mobile sensing, we also propose a scheme that employs the redundancy in security to reduce the communication cost of dealing with dynamic joins and leaves.

## CONCLUSION

The data aggregation scheme is designed for the mobile sensing applications. Data security and privacy preservation features are applied in the system. Our scheme has much lower communication overhead than existing work. The system utilizes the redundancy in security to reduce the communication cost for each join and leave.

## REFERENCES

[1]. Qinghua Li, Guohong Cao and Thomas F. La Porta, "Efficient and Privacy-Aware Data Aggregation in Mobile Sensing", IEEE Transactions On Dependable And Secure Computing, Vol. 11, No. 2, March/April 2014

[2]. Q. Li and G. Cao, "Providing Privacy-Aware Incentives for Mobile Sensing," Proc. IEEE PerCom, 2013.

[3]. Q. Li and G. Cao, "Efficient Privacy-Preserving Stream Aggregation in Mobile Sensing with Low Aggregation Error," Privacy Enhancing Technologies Symposium (PETS), 2013.

[4]. J. Hicks, N. Ramanathan, D. Kim, M. Hansen and D. Estrin, "AndWellness: An Open Mobile System for Activity and Experience Sampling," Proc. Wireless Health, pp. 34- 43, 2010.

[5]. N.D. Lane, M. Mohammod, E. Berke, T. Choudhury and A. Campbell, "Bewell: A Smartphone Application to Monitor, Model and Promote Wellbeing," Proc. Fifth Int'l ICST Conf. Pervasive Computing Technologies for Healthcare, 2011.

[6]. V. Rastogi and S. Nath, "Differentially Private Aggregation of Distributed Time-Series with Transformation and Encryption," Proc. ACM SIGMOD Int'l Conf. Management of Data, 2010.

[7]. E. Shi, T.-H.H. Chan, E. Rieffel, R. Chow and D. Song, "Privacy- Preserving Aggregation of Time-Series Data," Proc. Network and Distributed System Security Symp. (NDSS '11), 2011.

[8]. T.-H.H. Chan, E. Shi and D. Song, "Privacy-Preserving Stream Aggregation with Fault Tolerance," Proc. Sixth Int'l Conf. Financial Cryptography and Data Security (FC '12), 2012.

[9]. E.G. Rieffel, J. Biehl, W. van Melle and A.J. Lee, "Secured Histories: Computing Group Statistics on Encrypted Data While Preserving Individual Privacy,", 2010.

[10]. Q. Li and G. Cao, "Mitigating Routing Misbehavior in Disruption Tolerant Networks," IEEE Trans. Information Forensics and Security, vol. 7, no. 2, pp. 664 675, Apr. 2012.

[11]. MNDOLI, "Mnosha Permissible Exposure Limits," http:// www.dli.mn.gov/OSHA/PDF/pels.pdf, 2013.

[12]. Z. Erkin and G. Tsudik, "Private Computation of Spatial and Temporal Power Consumption with Smart Meters," Proc. Int'l Conf. Applied Cryptography and Network Security (ACNS '12), pp. 561-577, 2012.

[13]. G. Acs and C. Castelluccia, "I Have a Dream!: Differentially Private Smart Metering," Proc. 13th Int'l Conf. Information Hiding (IH '11), pp. 118-132, 2011.

[14]. M. Jawurek and F. Kerschbaum, "Fault-Tolerant Privacy- Preserving Statistics," Proc. 12th Privacy Enhancing Technologies Symp. (PETS '12), 2012.